

BIG DATA LAB

Learning component 2

Amith Kumar
1NT18IS027

GITHUB:

https://github.com/Amith027/BigData_lab_2021/tree/main/Bd_La2_Ex2

Exercise-I (MAP REDUCE)

Create a dataset in excel as .csv file and it should contain the following fields with at least 20 sample datasets in it.

Name	SSN	Salary	Address	Dname	Experience
Harsha	5000	30000	Bangalore	ISE	5

CSV FILE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Name	SSN	Salary	Address	Dname	Experience								
2	Harsha	9000	30000	Patna	Civil	3								
3	Amith	9001	30001	Ballari	ISE	2								
4	Kumar	9002	30002	Ballari	ISE	2								
5	Chinmaya	9003	30003	Mangaluru	ISE	1								
6	Krishna	9004	30004	Karkal	ISE	2								
7	Bharath	9005	30005	Ballari	ISE	1								
8	Gagan	9006	30006	Delhi	EEE	1								
9	Deepu	9007	30007	Delhi	EEE	3								
10	Smriti	9008	30008	Mumbai	CSE	3								
11	Hari	9009	30009	Hubli	CSE	2								
12	Shahsi	9010	30010	Bengaluru	CSE	2								
13	Manka	9011	30011	Shivmoga	ECE	1								
14	Nikhil	9012	30012	Shivmoga	ISE	2								
15	Prasanna	9013	30013	Vijayapura	MECH	3								
16	Aftab	9014	30014	Ballari	EEE	1								
17	Vaibhav	9015	30015	Dharwad	ECE	2								
18	Sujan	9016	30016	Mangaluru	ISE	3								
19	Muthu	9017	30017	Gulbarga	ECE	2								
20	Harish	9018	30018	Tumkur	CSE	1								
21	Arjun	9019	30019	Kolar	MECH	1								

Use the Hadoop MapReduce programming framework to come up with a Program which will take the data from this .csv file and computes the following.

1. Total number of employees who work in ISE department

JAR file execution

```
Activities Terminal Jul 10 00:01
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC: ~
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls
Found 2 items
drwxr-xr-x  - hadoop supergroup          0 2021-07-07 14:01 amith
-rw-r--r--  1 hadoop supergroup        680 2021-07-09 23:29 emp1 input.csv
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop jar /home/jayesh/hadoop-workspace/employee_1/EMP1.jar EmpISE emp1 input.csv emp1 output
2021-07-09 23:51:35,098 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 23:51:35,406 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 23:51:35,687 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-09 23:51:35,741 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625854542750_0002
2021-07-09 23:51:35,920 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-09 23:51:36,147 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 23:51:36,304 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-09 23:51:36,370 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-09 23:51:36,802 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 23:51:37,358 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-09 23:51:37,801 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625854542750_0002
2021-07-09 23:51:37,801 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 23:51:38,349 INFO conf.Configuration: resource-types.xml not found
2021-07-09 23:51:38,349 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-09 23:51:39,137 INFO impl.YarnClientImpl: Submitted application application_1625854542750_0002
2021-07-09 23:51:39,333 INFO mapreduce.Job: The url to track the job: http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625854542750_0002/
2021-07-09 23:51:39,336 INFO mapreduce.Job: Running job: job_1625854542750_0002
2021-07-09 23:52:07,409 INFO mapreduce.Job: Job job_1625854542750_0002 running in uber mode : false
2021-07-09 23:52:07,590 INFO mapreduce.Job:  map 0% reduce 0%
2021-07-09 23:52:44,920 INFO mapreduce.Job:  map 100% reduce 0%
2021-07-09 23:53:03,239 INFO mapreduce.Job:  map 100% reduce 100%
2021-07-09 23:53:05,263 INFO mapreduce.Job: Job job_1625854542750_0002 completed successfully
2021-07-09 23:53:05,443 INFO mapreduce.Job: Counters: 54
File System Counters
FILE: Number of bytes read=120
FILE: Number of bytes written=677735
FILE: Number of read operations=0
FILE: Number of large read operations=0
```

OUTPUT

```
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls
Found 3 items
drwxr-xr-x  - hadoop supergroup          0 2021-07-07 14:01 amith
-rw-r--r--  1 hadoop supergroup        680 2021-07-09 23:29 emp1 input.csv
drwxr-xr-x  - hadoop supergroup          0 2021-07-09 23:53 emp1 output
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls emp1 output
Found 2 items
-rw-r--r--  1 hadoop supergroup          0 2021-07-09 23:53 emp1 output/_SUCCESS
-rw-r--r--  1 hadoop supergroup         53 2021-07-09 23:53 emp1 output/part-00000
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -cat emp1 output/part-00000
2021-07-09 23:55:21,595 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees working in ISE Department :      7
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$
```

2. Total number of employees with experience = 2 years

JAR file execution

```
Activities Terminal Jul 10 00:36
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC: ~
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop jar /home/jayesh/hadoop-workspace/employee_1/emp1_exp.jar EmpExp emp1_input.csv emp1_output_exp
2021-07-10 00:33:48,564 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 00:33:49,226 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 00:33:49,892 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 00:33:50,017 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625854542750_0003
2021-07-10 00:33:50,332 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 00:33:50,660 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 00:33:50,870 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 00:33:50,970 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 00:33:51,402 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 00:33:51,859 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 00:33:51,946 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625854542750_0003
2021-07-10 00:33:51,946 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 00:33:52,574 INFO conf.Configuration: resource-types.xml not found
2021-07-10 00:33:52,575 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 00:33:52,888 INFO impl.YarnClientImpl: Submitted application application_1625854542750_0003
2021-07-10 00:33:53,314 INFO mapreduce.Job: The url to track the job: http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625854542750_0003/
2021-07-10 00:33:53,317 INFO mapreduce.Job: Running job: job_1625854542750_0003
2021-07-10 00:34:07,591 INFO mapreduce.Job: Job job_1625854542750_0003 running in uber mode : false
2021-07-10 00:34:07,610 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 00:34:33,585 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 00:35:00,156 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 00:35:04,197 INFO mapreduce.Job: Job job_1625854542750_0003 completed successfully
2021-07-10 00:35:04,388 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=6
  FILE: Number of bytes written=677528
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1218
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=11
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=45021
  Total time spent by all reduces in occupied slots (ms)=21027
  Total time spent by all map tasks (ms)=45021
  Total time spent by all reduce tasks (ms)=21027
  Total vcore-milliseconds taken by all map tasks=45021
  Total vcore-milliseconds taken by all reduce tasks=21027
```

OUTPUT

```
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls
Found 4 items
drwxr-xr-x - hadoop supergroup 0 2021-07-07 14:01 amith
-rw-r--r-- 1 hadoop supergroup 680 2021-07-09 23:29 emp1_input.csv
drwxr-xr-x - hadoop supergroup 0 2021-07-09 23:53 emp1_output
drwxr-xr-x - hadoop supergroup 0 2021-07-10 00:48 emp1_output_exp
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls emp1_output_exp
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-07-10 00:48 emp1_output_exp/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 56 2021-07-10 00:48 emp1_output_exp/part-00000
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -cat emp1_output_exp/part-00000
2021-07-10 00:49:46,854 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees having 2 years of experience : 8
hadoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$
```


3. Count the number of employees who live in Bengaluru.

JAR file execution

```
Activities Terminal Jul 10 01:17
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC: ~
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop jar /home/jayesh/hadoop-workspace/employee_1/emp1_add.jar ExpAdd emp1_input.csv emp1_output_add
2021-07-10 01:10:52,848 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 01:10:53,365 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 01:10:53,820 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 01:10:53,960 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625854542750_0007
2021-07-10 01:10:54,333 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 01:10:55,086 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 01:10:55,356 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 01:10:56,676 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 01:10:57,375 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 01:10:57,855 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-10 01:10:58,517 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625854542750_0007
2021-07-10 01:10:58,518 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 01:10:59,372 INFO conf.Configuration: resource-types.xml not found
2021-07-10 01:10:59,373 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 01:10:59,739 INFO impl.YarnClientImpl: Submitted application application_1625854542750_0007
2021-07-10 01:10:59,848 INFO mapreduce.Job: The url to track the job: http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625854542750_0007/
2021-07-10 01:10:59,850 INFO mapreduce.Job: Running job: job_1625854542750_0007
2021-07-10 01:11:09,954 INFO mapreduce.Job: Job job_1625854542750_0007 running in uber mode : false
2021-07-10 01:11:09,956 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 01:11:21,286 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 01:11:41,665 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 01:11:46,713 INFO mapreduce.Job: Job job_1625854542750_0007 completed successfully
2021-07-10 01:11:46,878 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=58
  FILE: Number of bytes written=677608
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1218
  HDFS: Number of bytes written=48
  HDFS: Number of read operations=11
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=17310
  Total time spent by all reduces in occupied slots (ms)=16430
  Total time spent by all map tasks (ms)=17310
  Total time spent by all reduce tasks (ms)=16430
  Total vcore-milliseconds taken by all map tasks=17310
  Total vcore-milliseconds taken by all reduce tasks=16430
```

OUTPUT

```
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls
Found 5 items
drwxr-xr-x - hdoop supergroup 0 2021-07-07 14:01 amith
-rw-r--r-- 1 hdoop supergroup 680 2021-07-09 23:29 emp1_input.csv
drwxr-xr-x - hdoop supergroup 0 2021-07-09 23:53 emp1_output
drwxr-xr-x - hdoop supergroup 0 2021-07-10 01:11 emp1_output_add
drwxr-xr-x - hdoop supergroup 0 2021-07-10 00:48 emp1_output_exp
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -ls emp1_output_add
Found 2 items
-rw-r--r-- 1 hdoop supergroup 0 2021-07-10 01:11 emp1_output_add/_SUCCESS
-rw-r--r-- 1 hdoop supergroup 48 2021-07-10 01:11 emp1_output_add/part-00000
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$ hadoop fs -cat emp1_output_add/part-00000
2021-07-10 01:14:50,692 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees working in Bengaluru : 1
hdoop@jayesh-HP-PAVILION-G6-NOTEBOOK-PC:~$
```

PRGM1 Employees in ISE dept.

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;
```

```
public class EmpISE {
    //MAPPER CODE
```

```
    public static class Map extends MapReduceBase implements
    Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();
```

```
    public void map(LongWritable key, Text value, OutputCollector<Text,
    IntWritable> output, Reporter reporter) throws IOException {
        String data = value.toString();
        String[] Ecount= data.split(",");
```

```
        if (Ecount[4].equals("ISE")) {
            output.collect(new Text("Total no.of employees working in ISE
    Department : "), one);
        }
    }
}
```

```
//REDUCER CODE
```

```
public static class Reduce extends MapReduceBase implements
Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterator<IntWritable> values,
    OutputCollector<Text, IntWritable> output, Reporter reporter)
    throws IOException { //{little: {1,1}}
        int val = 0 ;
```

```

while(values.hasNext()) {
    val += values.next().get();
}
output.collect(key, new IntWritable(val));
}
}

```

//DRIVER CODE

```

public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(EmpISE.class);
    conf.setJobName("Total no.of employees working in ISE
Department");
    conf.setOutputKeyClass(Text.class);
    conf.setOutputValueClass(IntWritable.class);
    conf.setMapperClass(Map.class);
    conf.setCombinerClass(Reduce.class);
    conf.setReducerClass(Reduce.class);
    conf.setInputFormat(TextInputFormat.class);
    conf.setOutputFormat(TextOutputFormat.class);
    FileInputFormat.setInputPaths(conf, new Path(args[0]));
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));
    JobClient.runJob(conf);
}
}

```

PRGM2 Employees with experience more than 2 yrs.

```

import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;

```

```

public class EmpExp {
    //Mapper Code

```

```

public static class Map extends MapReduceBase implements
Mapper<LongWritable, Text, Text, IntWritable> {
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();

    public void map(LongWritable key, Text value,
OutputCollector<Text, IntWritable> output, Reporter reporter)
throws IOException {
        String data = value.toString();
        String[] Ecount= data.split(",");

        if (Ecount[5].equals("2")) {
            output.collect(new Text("Total no.of employees having 2
years of experience : "), one);
        }

    }
}

```

//REDUCER CODE

```

public static class Reduce extends MapReduceBase implements
Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterator<IntWritable> values,
OutputCollector<Text, IntWritable> output, Reporter reporter)
throws IOException { //{little: {1,1}}
        int val = 0 ;
        while(values.hasNext()) {
            val += values.next().get();
        }
        output.collect(key, new IntWritable(val));
    }
}

```

//DRIVER CODE

```

public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(EmpExp.class);
    conf.setJobName("Total no.of employees having 5 years of
experience");
    conf.setOutputKeyClass(Text.class);
}

```

```

        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setCombinerClass(Reduce.class);
        conf.setReducerClass(Reduce.class);
        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        JobClient.runJob(conf);
    }
}

```

PRGM3 Employees in Bengaluru.

```

import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;

public class ExpAdd {
    //Mapper Code
    public static class Map extends MapReduceBase implements
    Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(LongWritable key, Text value,
        OutputCollector<Text, IntWritable> output, Reporter reporter)
        throws IOException {
            String data = value.toString();
            String[] Ecount= data.split(",");

            if (Ecount[3].equals("Bengaluru")) {

```



```
        output.collect(new Text("Total no.of employees working in  
Bengaluru : "), one);  
    }  
  
}
```

//REDUCER CODE

```
public static class Reduce extends MapReduceBase implements  
Reducer<Text, IntWritable, Text, IntWritable> {  
    public void reduce(Text key, Iterator<IntWritable> values,  
OutputCollector<Text, IntWritable> output, Reporter reporter)  
throws IOException { //{little: {1,1}}  
        int val = 0 ;  
        while(values.hasNext()) {  
            val += values.next().get();  
        }  
        output.collect(key, new IntWritable(val));  
    }  
}
```

//DRIVER CODE

```
public static void main(String[] args) throws Exception {  
    JobConf conf = new JobConf(EmpExp.class);  
    conf.setJobName("Total no.of employees working in  
Bengaluru");  
    conf.setOutputKeyClass(Text.class);  
    conf.setOutputValueClass(IntWritable.class);  
    conf.setMapperClass(Map.class);  
    conf.setCombinerClass(Reduce.class);  
    conf.setReducerClass(Reduce.class);  
    conf.setInputFormat(TextInputFormat.class);  
    conf.setOutputFormat(TextOutputFormat.class);  
    FileInputFormat.setInputPaths(conf, new Path(args[0]));  
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));  
    JobClient.runJob(conf);  
}
```

Exercise-II

Use the above dataset in .csv file and create a database called as EmployeeDB. Create a table under the database called as Employee using HIVEQL. The table fields are same, that is,

Name	SSN	Salary	Address	Dname	Experience
Harsha	5000	30000	Bangalore	ISE	5

1) Create database employee db and a table in it with named employee

```
hive> show databases;
OK
default
dummy
Time taken: 0.037 seconds, Fetched: 2 row(s)
hive> create database Employee;
OK
Time taken: 0.142 seconds
hive> show databases;
OK
default
dummy
employee
Time taken: 0.032 seconds, Fetched: 3 row(s)
hive> use Employee;
OK
Time taken: 0.035 seconds
hive> show tables;
OK
Time taken: 0.04 seconds
hive> create table employee(Name string, SSN int, Salary int, Address string, Dname string, Experience int);
OK
Time taken: 0.183 seconds
hive> desc employee;
OK
name                string
ssn                  int
salary               int
address              string
dname                string
experience            int
Time taken: 0.094 seconds, Fetched: 6 row(s)
hive> █
```

2. Insert 5 records using INSERT command

```
hive> insert into employee values('Amith',5000,50000,'Bengaluru','ISE',6),
> ('Chinmaya',5001,45000,'Mangaluru','ISE',4),
> ('Bharath',5002,60000,'Bengaluru','CSE',5),
> ('Vaibhav',5003,120000,'Ballari','ECE',2),
> ('Aftab',5004,30000,'Ballari','EEE',3);
Query ID = hdoop_20210710144254_4c833683-a9f3-4559-a720-cd6288bee39d
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0001, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0001/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 14:43:25,533 Stage-1 map = 0%, reduce = 0%
2021-07-10 14:43:35,110 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.2 sec
2021-07-10 14:43:42,507 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.55 sec
MapReduce Total cumulative CPU time: 7 seconds 550 msec
Ended Job = job_1625904496337_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employee.db/employee/.hive-staging_hive_2021-07-10_14-42-54_203_5773417581175335080-1/-ext-10000
Loading data to table employee.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.55 sec HDFS Read: 20975 HDFS Write: 663 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 550 msec
OK
Time taken: 51.775 seconds
hive> select * from employee;
OK
Amith 5000 50000 Bengaluru ISE 6
Chinmaya 5001 45000 Mangaluru ISE 4
Bharath 5002 60000 Bengaluru CSE 5
Vaibhav 5003 120000 Ballari ECE 2
Aftab 5004 30000 Ballari EEE 3
Time taken: 0.474 seconds, Fetched: 5 row(s)
hive>
```

3. Demonstrate the Alter command for the following cases, Rename the table name to “Emp”. Rename the column name “Dname” to “Dept_name”

```
> show tables;
OK
employee
Time taken: 0.038 seconds, Fetched: 1 row(s)
hive> alter table employee rename to Emp;
OK
Time taken: 0.384 seconds
hive> show tables;
OK
emp
Time taken: 0.037 seconds, Fetched: 1 row(s)
hive> alter table emp change dname dept_name string;
OK
Time taken: 0.216 seconds
hive> desc emp;
OK
name                string
ssn                  int
salary               int
address              string
dept_name            string
experience            int
Time taken: 0.062 seconds, Fetched: 6 row(s)
hive>
```

4. Retrieve all the employees who's salary is not less than 50000.

```
hive> select * from emp where salary>=50000;
OK
Amith 5000 50000 Bengaluru ISE 6
Bharath 5002 60000 Bengaluru CSE 5
Vaibhav 5003 120000 Ballari ECE 2
Time taken: 0.416 seconds, Fetched: 3 row(s)
hive>
```

5. Extract all employees who live in Bangalore but having less than 5 years of experience.

```
hive> select * from emp where address='Bengaluru' and experience>5;
OK
Amith 5000 50000 Bengaluru ISE 6
Time taken: 0.354 seconds, Fetched: 1 row(s)
hive>
```

6. Create separate view containing Name, Dept_name of employees

```
hive> create view details as (select name, dept_name from emp);
OK
Time taken: 0.589 seconds
hive> select * from details;
OK
Amith ISE
Chinmaya ISE
Bharath CSE
Vaibhav ECE
Aftab EEE
Time taken: 0.291 seconds, Fetched: 5 row(s)
hive>
```


7. Display Name and SSN and use group by SSN and order by Name

```
hive> select name,ssn from emp group by ssn,name order by name;
Query ID = hdoop_20210710152858_d8c834be-84fa-4a99-be9a-0accbb94a717
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0002, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0002/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 15:29:10,978 Stage-1 map = 0%, reduce = 0%
2021-07-10 15:29:18,511 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.15 sec
2021-07-10 15:29:25,852 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.24 sec
MapReduce Total cumulative CPU time: 6 seconds 240 msec
Ended Job = job_1625904496337_0002
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0003, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0003/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-10 15:29:42,316 Stage-2 map = 0%, reduce = 0%
2021-07-10 15:29:48,562 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.58 sec
2021-07-10 15:29:55,887 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.89 sec
MapReduce Total cumulative CPU time: 5 seconds 890 msec
Ended Job = job_1625904496337_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.24 sec HDFS Read: 11946 HDFS Write: 233 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.89 sec HDFS Read: 7643 HDFS Write: 209 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 130 msec
OK
Aftab 5004
Amith 5000
Bharath 5002
Chinnmaya 5001
Vaibhav 5003
Time taken: 59.754 seconds, Fetched: 5 row(s)
hive> █
```

8. Retrieve Maximum salary, minimum salary and Average salary of the employees

```
hive> select min(salary),max(salary),avg(salary) from emp;
Query ID = hdoop_20210710153201_b0db9571-560b-46c0-b610-04e43d6fc248
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0004, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0004/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 15:32:13,827 Stage-1 map = 0%, reduce = 0%
2021-07-10 15:32:21,157 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.93 sec
2021-07-10 15:32:29,464 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.52 sec
MapReduce Total cumulative CPU time: 8 seconds 520 msec
Ended Job = job_1625904496337_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.52 sec HDFS Read: 17232 HDFS Write: 120 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 520 msec
OK
30000 120000 61000.0
Time taken: 28.718 seconds, Fetched: 1 row(s)
hive> █
```

9. Create Another table called Department with the following fields (Dname = Dept_name and perform the following joins (outer, left outer, right outer) over Dname

```
hive> create table department(dno int, dname string);
OK
Time taken: 0.217 seconds
hive> show tables;
OK
department
details
emp
Time taken: 0.037 seconds, Fetched: 3 row(s)
hive> desc department;
OK
dno                int
dname              string
Time taken: 0.072 seconds, Fetched: 2 row(s)
hive> insert into department values(6,'ISE'),(5,'ECE'),(7,'EEE');
Query ID = hdoop_20210710153505_c48f0fb6-2a8d-4095-b947-79745aa83721
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625904496337_0005, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0005/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 15:35:18,952 Stage-1 map = 0%, reduce = 0%
2021-07-10 15:35:26,320 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.16 sec
2021-07-10 15:35:34,644 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.69 sec
MapReduce Total cumulative CPU time: 6 seconds 690 msec
Ended Job = job_1625904496337_0005
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employee.db/department/.hive-staging_hive_2021-07-10_15-35-05_976_3910694052797637697-1/-ext-10000
Loading data to table employee.department
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.69 sec HDFS Read: 15614 HDFS Write: 280 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 690 msec
OK
Time taken: 31.337 seconds
hive> select * from department;
OK
6      ISE
5      ECE
7      EEE
Time taken: 0.22 seconds, Fetched: 3 row(s)
hive>
```

10. Left outer Join

```
hive> select * from emp left outer join department where dname=dept_name;
Query ID = hdoop_20210710153700_fs4cdda9-22b0-4a41-8d53-bd47ac0f67a4
Total jobs = 1
```

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

2021-07-10 15:37:17 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/hive/java/hadoop/310c2f3c-bd0f-43a4-9980-56c7c615f0ef/hive_2021-07-10_15-37-00_765_7966489225870597337-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable2021-07-10 15:37:17 Uploaded 1 File to: file:/tmp/hive/java/hadoop/310c2f3c-bd0f-43a4-9980-56c7c615f0ef/hive_2021-07-10_15-37-00_765_7966489225870597337-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (329 bytes)

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1625904496337_0006, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0006/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0006

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2021-07-10 15:37:30,959 Stage-3 map = 0%, reduce = 0%

2021-07-10 15:37:39,311 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.17 sec

MapReduce Total cumulative CPU time: 5 seconds 170 msec

Ended Job = job_1625904496337_0006

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 5.17 sec HDFS Read: 10418 HDFS Write: 293 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 170 msec

OK

Amith	5000	50000	Bengaluru	ISE	6	6	ISE
Chinnaya	5001	45000	Mangaluru	ISE	4	6	ISE
Vaibhav	5003	120000	Ballari	ECE	2	5	ECE
Aftab	5004	30000	Ballari	EEE	3	7	EEE

Time taken: 40.776 seconds, Fetched: 4 row(s)

hive>

11. Right outer join

```
hive> select * from emp right outer join department where dname=dept_name;
Query ID = hdoop_20210710153825_affe9cc7-d50c-462a-8535-4e01f72d5b23
Total jobs = 1
```

2021-07-10 15:38:35 Starting to launch local task to process map join; maxinum memory = 239075328

2021-07-10 15:38:38 Uploaded 1 File to: file:/tmp/hive/java/hadoop/310c2f3c-bd0f-43a4-9980-56c7c615f0ef/hive_2021-07-10_15-38-25_780_8433326092268546618-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile11--.hashtable (329 bytes)

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1625904496337_0007, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8088/proxy/application_1625904496337_0007/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0007

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2021-07-10 15:38:50,761 Stage-3 map = 0%, reduce = 0%

2021-07-10 15:38:59,173 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.2 sec

MapReduce Total cumulative CPU time: 5 seconds 200 msec

Ended Job = job_1625904496337_0007

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 5.2 sec HDFS Read: 10431 HDFS Write: 293 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 200 msec

OK

Amith	5000	50000	Bengaluru	ISE	6	6	ISE
Chinnaya	5001	45000	Mangaluru	ISE	4	6	ISE
Vaibhav	5003	120000	Ballari	ECE	2	5	ECE
Aftab	5004	30000	Ballari	EEE	3	7	EEE

Time taken: 34.524 seconds, Fetched: 4 row(s)

hive>

12. Full Outer join

```
hive> select * from emp full outer join department where dname=dept_name;
Query ID = hdoop_20210710153949_9556189d-8672-4ade-acbf-83db3feba343
Total jobs = 1

SLF4J: Found binding in [jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2021-07-10 15:40:00 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/hive/java/hadoop/310c2f3c-bd0f-43a4-9980-56c7c615f0ef/hive_2021-07-10_15-39-49_119_2074615144985493783-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile21--hashtable2021-07-10 15:40:00 Uploaded 1 File to: file:/tmp/hive/java/hadoop/310c2f3c-bd0f-43a4-9980-56c7c615f0ef/hive_2021-07-10_15-39-49_119_2074615144985493783-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile21--hashtable (329 bytes)
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1625904496337_0000, Tracking URL = http://jayesh-HP-PAVILION-G6-NOTEBOOK-PC:8080/proxy/application_1625904496337_0000/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625904496337_0000
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-10 15:40:12,100 Stage-3 map = 0%, reduce = 0%
2021-07-10 15:40:20,445 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.3 sec
MapReduce Total cumulative CPU time: 5 seconds 300 msec
Ended Job = job_1625904496337_0000
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 5.3 sec HDFS Read: 10431 HDFS Write: 293 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 300 msec
OK
Amith 5000 50000 Bengaluru ISE 6 6 ISE
Chinnaya 5001 45000 Mangaluru ISE 4 6 ISE
Vaibhav 5003 120000 Ballari ECE 2 5 ECE
Aftab 5004 30000 Ballari EEE 3 7 EEE
Time taken: 32.49 seconds, Fetched: 4 row(s)
hive>
```

CODE:

```
hive> show databases;
```

```
OK
```

```
default
```

```
Time taken: 0.403 seconds, Fetched: 1 row(s) hive> create database employee;
```

```
OK
```

```
Time taken: 0.153 seconds
```

```
hive> use employee;
```

```
OK
```

```
Time taken: 0.036 seconds
```

```
hive> show tables;
```

```
OK
```

```
Time taken: 0.038 seconds
```

```
hive> create table employee(Name string,SSN int,Address string,Dname
```

```
String,Experience int);
```

```
OK
```

```
Time taken: 0.443 seconds
```

```
hive> show tables;
```

```
OK
```

```
employee
```

```
Time taken: 0.028 seconds, Fetched: 1 row(s)
```

```
hive> desc employee;
```

```
OK
```

```
name
```

```
ssn
```

```
address
```

```
dname
```

```
experience
```

```
Time taken: 0.132 seconds, Fetched: 5 row(s)
```

```
hive> insert into employee values('chinmaya',5000,12000,'Bangalore','ISE',6),
```

```
> ('Shashank',5001,42000,'Shimoga','EC',4), >
```

```
('Amith',5002,22000,'Bangalore','ISE',5),
```

```
> ('Krushik',5003,50000,'Mangalore','EEE',2), >
```

```
('Nikhil',5004,52000,'Ballary','AERO',3);
```

```
FAILED: SemanticException [Error 10044]: Line 1:12 Cannot insert into target table  
because column number/types are different 'employee': Table insclause-0 has 5  
columns, but query has 6 columns.
```

```
hive> drop table employee;
```

```
OK
```

```
Time taken: 0.478 seconds hive> show tables;
```

```
OK
```

```
Time taken: 0.025 seconds
```

```
string int string string int
```

```
hive> create table employee(Name string,SSN int,Salary int,Address string,Dname  
String,Experience int);
```

```
OK
```

```
Time taken: 0.109 seconds
```

```
hive> insert into employee values('chinmaya',5000,12000,'Bangalore','ISE',6), >
```

```
('Shashank',5001,42000,'Shimoga','EC',4),
```

```
> ('Amith',5002,22000,'Bangalore','ISE',5),
```

```
> ('Krushik',5003,50000,'Mangalore','EEE',2),
```

```
> ('Nikhil',5004,52000,'Ballary','AERO',3);
Query ID = hdoop_20210710120100_c46890b9-092e-45ec-84e1-4315c11e11c4
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
set hive.exec.reducers.bytes.per.reducer=<number> In order to limit the maximum
number of reducers:
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapreduce.job.reduces=<number>
Starting Job = job_1625897919713_0001, Tracking URL = http://Chinmayakb:8088/
proxy/application_1625897919713_0001/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill
job_1625897919713_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-10 12:01:09,670 Stage-1 map = 0%, reduce = 0%
2021-07-10 12:01:15,889 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.19
sec
2021-07-10 12:01:21,007 Stage-1 map = 100%, reduce = 100%, Cumulative CPU
3.75 sec
MapReduce Total cumulative CPU time: 3 seconds 750 msec
Ended Job = job_1625897919713_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory
hdfs://127.0.0.1:9000/user/hive/warehouse/employee.db/employee/.hive-
staging_hive_2021-07-10_12-01-00_533_3418192329285162114-1/-ext-10000
Loading data to table employee.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.75 sec HDFS Read: 20983
HDFS Write: 671 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 750 msec
OK
Time taken: 23.141 seconds
hive> select * from employee;
```

OK chinmaya Shashank Amith Krushik Nikhil Time taken:

5000 12000Bangalore ISE 6

5001 42000Shimoga EC4

5002 22000Bangalore ISE 5

5003 50000Mangalore EEE 2

5004 52000Ballary AERO 3 0.206 seconds, Fetched: 5 row(s)

> show tables;

OK

employee

Time taken: 0.025 seconds, Fetched: 1 row(s) hive> alter table employee rename to EMP; OK

Time taken: 0.153 seconds hive> show tables;

OK

emp

Time taken: 0.028 seconds, Fetched: 1 row(s)

hive> alter table emp change dname dept_name string;

OK

Time taken: 0.121 seconds

hive> desc emp;

OK

name

ssn

salary

address

dept_name

experience

Time taken:

hive> select

OK

Krushik

Nikhil

Time taken:

hive> select

OK

Time taken:

hive> select

OK

chinmaya

Time taken:

hive> create view details as (select name,dept_name from emp); OK

string int int string string int

0.04 seconds, Fetched: 6 row(s)

* from emp where salary>=50000;

5003 50000Mangalore EEE 2

5004 52000Ballary AERO 3

0.212 seconds, Fetched: 2 row(s)

* from emp where address ='Bangalore' and experience<5;

0.161 seconds

* from emp where address ='Bangalore' and experience>5;

5000 12000Bangalore ISE 6 0.136 seconds, Fetched: 1 row(s)

0.159 seconds * from details;

Time taken:

hive> select

OK

chinmaya

Shashank

Amith

Krushik

Nikhil

Time taken:

hive> select

Query ID = hdoop_20210710120631_8239df8f-b6af-4117-bbe4-bb1592299fbd Total jobs = 2

Launching Job 1 out of 2

Number of reduce tasks not specified. Estimated from input data size: 1 In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number> In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1625897919713_0002, Tracking URL = <http://Chinmayakb:8088/>

proxy/application_1625897919713_0002/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill

job_1625897919713_0002

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-07-10 12:06:38,147 Stage-1 map = 0%, reduce = 0%

2021-07-10 12:06:43,301 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec

2021-07-10 12:06:49,510 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.86 sec

MapReduce Total cumulative CPU time: 2 seconds 860 msec

Ended Job = job_1625897919713_0002

Launching Job 2 out of 2

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number> In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1625897919713_0003, Tracking URL = http://Chinmayakb:8088/proxy/application_1625897919713_0003/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill

job_1625897919713_0003

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

ISE

EC

ISE

EEE

AERO

0.166 seconds, Fetched: 5 row(s)

name,ssn from emp group by ssn,name order by name;

2021-07-10 12:07:00,442 Stage-2 map = 0%, reduce = 0%

2021-07-10 12:07:10,726 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.12 sec

2021-07-10 12:07:15,885 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.99 sec

MapReduce Total cumulative CPU time: 3 seconds 990 msec

Ended Job = job_1625897919713_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.86 sec HDFS Read: 11947
HDFS Write: 235 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.99 sec HDFS Read: 7645
HDFS Write: 211 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 850 msec

OK

Amith 5002

Krushik 5003

Nikhil 5004

Shashank 5001

chinmaya 5000

Time taken: 50.715 seconds, Fetched: 5 row(s)

hive> select min(salary),max(salary),avg(salary) from emp;

Query ID = hdoop_20210710121111_1f39f9c4-8e66-4baa-8c5a-b71a6c33f060

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number> In order to limit the maximum
number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1625897919713_0004, Tracking URL = [http://Chinmayakb:8088/
proxy/application_1625897919713_0004/](http://Chinmayakb:8088/proxy/application_1625897919713_0004/)

Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill

job_1625897919713_0004

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-07-10 12:11:17,569 Stage-1 map = 0%, reduce = 0%

2021-07-10 12:11:21,727 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.67
sec

2021-07-10 12:11:27,908 Stage-1 map = 100%, reduce = 100%, Cumulative CPU
4.43 sec

MapReduce Total cumulative CPU time: 4 seconds 430 msec

Ended Job = job_1625897919713_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.43 sec HDFS Read: 17374
HDFS Write: 119 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 430 msec
OK

120005200035600.0

Time taken: 17.583 seconds, Fetched: 1 row(s)

hive> create table department(dno int,dname string); OK

Time taken: 0.12 seconds

hive> show tables;

OK

department

details

emp

Time taken: 0.03 seconds, Fetched: 3 row(s)

hive> desc department;

OK

dno int

dname string

Time taken: 0.027 seconds, Fetched: 2 row(s)

hive> insert into department values(6,'ISE'),(5,'EC'),(7,'AERO');

Query ID = hdoop_20210710121402_e316535a-0a4d-4736-adf2-5119c7f377c0 Total
jobs = 3

Launching Job 1 out of 3

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number> In order to limit the maximum
number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1625897919713_0005, Tracking URL = [http://Chinmayakb:8088/
proxy/application_1625897919713_0005/](http://Chinmayakb:8088/proxy/application_1625897919713_0005/)

Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill

job_1625897919713_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-07-10 12:14:07,651 Stage-1 map = 0%, reduce = 0%

2021-07-10 12:14:12,792 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.68

sec

2021-07-10 12:14:16,914 Stage-1 map = 100%, reduce = 100%, Cumulative CPU
4.34 sec

MapReduce Total cumulative CPU time: 4 seconds 340 msec

Ended Job = job_1625897919713_0005

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to directory

hdfs://127.0.0.1:9000/user/hive/warehouse/employee.db/department/.hive-
staging_hive_2021-07-10_12-14-02_449_9028547012151235819-1/-ext-10000

Loading data to table employee.department

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.34 sec HDFS Read: 15628
HDFS Write: 280 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 340 msec

OK

Time taken: 15.809 seconds

hive> select * from department;

OK

6 ISE

5 EC

7 AERO

Time taken: 0.111 seconds, Fetched: 3 row(s)

hive> select * from emp left outer join department where dname=dept_name; Query
ID = hdoop_20210710121555_13bc8842-0de7-4185-becf-f28133639a3d Total jobs =
1

SLF4J: Found binding in

[jar:file:/home/hdoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

2021-07-10 12:16:01 Dump the side-table for tag: 1 with group count: 3 into file:
file:/tmp/hive/java/hdoop/f29000d8-b5cb-46b3-be29-729ff8a6392e/hive_2021-07-
10_12-15-55_658_5060518414468933463-1/-local-10004/HashTable-Stage-3/

MapJoin-mapfile01--.hashtable2021-07-10 12:16:01 Uploaded 1 File to:

file:/tmp/hive/java/hdoop/f29000d8-b5cb-46b3-be29-729ff8a6392e/hive_2021-07-
10_12-15-55_658_5060518414468933463-1/-local-10004/HashTable-Stage-3/

MapJoin-mapfile01--.hashtable (329 bytes)

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1625897919713_0006, Tracking URL = http://Chinmayakb:8088/proxy/application_1625897919713_0006/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill
job_1625897919713_0006

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2021-07-10 12:16:08,263 Stage-3 map = 0%, reduce = 0%

2021-07-10 12:16:12,381 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.45
sec

MapReduce Total cumulative CPU time: 2 seconds 450 msec

Ended Job = job_1625897919713_0006

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 2.45 sec HDFS Read: 10419 HDFS Write:
294 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 450 msec

OK

chinmaya 5000 12000Bangalore ISE 6 6 ISE

Shashank

Amith

Nikhil

Time taken: 24.036 seconds, Fetched: 4 row(s)

hive> select * from emp right outer join department where dname=dept_name; Query
ID = hdoop_20210710121645_a828ba60-0338-48e8-9029-88773660df32 Total jobs
= 1

SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-
impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in

[jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-
1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

2021-07-10 12:16:50 Dump the side-table for tag: 1 with group count: 3 into file:
file:/tmp/hive/java/hadoop/f29000d8-b5cb-46b3-be29-729ff8a6392e/hive_2021-07-

10_12-16-45_561_8658834088026560490-1/-local-10004/HashTable-Stage-3/
MapJoin-mapfile11--.hashtable

2021-07-10 12:16:50 End of local task; Time Taken: 1.151 sec. Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1625897919713_0007, Tracking URL = http://Chinmayakb:8088/proxy/application_1625897919713_0007/

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill
job_1625897919713_0007

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2021-07-10 12:16:57,442 Stage-3 map = 0%, reduce = 0%

2021-07-10 12:17:02,587 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.5 sec

MapReduce Total cumulative CPU time: 2 seconds 500 msec

Ended Job = job_1625897919713_0007

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 2.5 sec HDFS Read: 10432 HDFS Write: 294 SUCCESS

5001 42000Shimoga EC 4 5002 22000Bangalore ISE 5 5004 52000Ballary AERO

5 EC

6 ISE

3 7 AERO

Total MapReduce CPU Time Spent: 2 seconds 500 msec OK

chinmaya

Shashank

Amith

Nikhil

Time taken: 18.452 seconds, Fetched: 4 row(s)

hive> select * from emp full outer join department where dname=dept_name; Query ID = hdoop_20210710121720_f5c9d77c-9493-43c6-8f3f-4fa714303d1e Total jobs = 1

SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]

2021-07-10 12:17:25 Dump the side-table for tag: 1 with group count: 3 into file:

file:/tmp/hive/java/hadoop/f29000d8-b5cb-46b3-be29-729ff8a6392e/hive_2021-07-10_12-17-20_112_5810221896728672370-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile21--.hashtable2021-07-10 12:17:25 Uploaded 1 File to:
file:/tmp/hive/java/hadoop/f29000d8-b5cb-46b3-be29-729ff8a6392e/hive_2021-07-10_12-17-20_112_5810221896728672370-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile21--.hashtable (329 bytes)

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1625897919713_0008, Tracking URL = http://Chinmayakb:8088/proxy/application_1625897919713_0008/

5000 12000Bangalore ISE 6 5001 42000Shimoga EC 4 5002 22000Bangalore ISE 5
5004 52000Ballary AERO

6 ISE

5.5 EC

6.6 ISE

3 7 AERO

Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill
job_1625897919713_0008

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0

2021-07-10 12:17:31,105 Stage-3 map = 0%, reduce = 0%

2021-07-10 12:17:36,203 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.61
sec

MapReduce Total cumulative CPU time: 2 seconds 610 msec

Ended Job = job_1625897919713_0008

MapReduce Jobs Launched:

Stage-Stage-3: Map: 1 Cumulative CPU: 2.61 sec HDFS Read: 10432 HDFS Write:
294 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 610 msec

OK

chinmaya

Shashank

Amith

Nikhil

Time taken: 17.194 seconds, Fetched: 4 row(s)
hive>